

Segmentasi Perilaku Pembelian Pelanggan Berdasarkan Model RFM dengan Metode K-Means

Siti Monalisa

Program Studi Sistem Informasi

Universitas Islam Negeri Sultan Syarif Kasim Riau

siti.monalisa@uin-suska.ac.id

Abstract

This study aims to measure customer purchase behavior by using RFM model with K-Means, silhouette Index method. Using historical data by Cendana Perdana Perkasa Distributor to data this study. The company has never to measured customer buy behavior by this model. The results of this study showed that there 3 cluster with 0,589 silhouette Index. Based on the highest silhouette Index then the cluster is divided into 3 segment with some segments 1 is 57 customers, segment 2 is 257 customers and segment 3 is 8 customers. The best purchase value is segment 3 with symbolized by $R\downarrow F\uparrow M\uparrow$ that means this segment has high frequency and monetary as well as low regency. The segment has potential customers for the future. The worst purchase value is segment 1 with symbolised by $R\uparrow F\downarrow M\downarrow$ that means this segment has low-frequency and monetary as well as high regency. The segment has an unclear or uncertain customer will be a potential customer.

Keywords: Customer Purchase Behavior, K-Means, RFM and Silhouette Index.

1. PENDAHULUAN

Saat ini, perusahaan yang sukses adalah perusahaan yang dapat merespon perubahan dan peluang pasar dengan cepat dan fleksibel dengan cara memanfaatkan data dan informasi secara efektif dan efisien. Data dan informasi perusahaan bisa berasal dari sumber internal maupun eksternal. Salah satu sumber data internal perusahaan yang bisa dimanfaatkan adalah data pelanggan. Data pelanggan bisa dimanfaatkan untuk melakukan strategi pemasaran produk dan jasa dengan tepat ke masing-masing segmen pelanggan. Segmentasi pelanggan merupakan elemen yang digunakan untuk mengidentifikasi pelanggan dalam fase pertama CRM yaitu akuisisi pelanggan [6]. Menurut [1] segmentasi pelanggan adalah praktik mempartisi pelanggan menjadi himpunan bagian yang homogen sehingga masing-masing subset dapat ditangani sebagai khalayak pemasaran yang unik. Segmentasi pelanggan digunakan untuk mengukur nilai pelanggan (*Customer Value*) sehingga perusahaan bisa menentukan pelanggan mana yang memberikan keuntungan yang besar dan mana pelanggan yang tidak memberikan keuntungan. Hal ini dibedakan karena tidak semua pelanggan dilayani dan dipertahankan oleh perusahaan. Jika pelanggan tersebut tidak memberikan kontribusi berupa keuntungan kepada perusahaan maka perusahaan tidak perlu memberikan perhatian dan

Customer Value telah dipelajari dengan nama LTV (Lifetime Value), CLV (Customer Lifetime Value), CE (Customer Equity) dan CP (Customer Profitability) [5]. CLV telah digunakan oleh He dan Li (2016) dalam mengukur pembelian pelanggan. Analisis CLV ini berguna dalam memprediksi total pendapatan bersih perusahaan yang dapat diharapkan yang berasal dari pelanggan. Kriteria CLV bisa menggunakan model RFM (**Recency, Frequency dan Monetary**). Model RFM ini diperkenalkan pertama kali oleh Arthur Huges dan telah banyak digunakan oleh industri manufaktur, retailer dan industri jasa. Model RFM secara efektif mencapai *Customer Relationship Management* karena model ini merupakan salah satu sarana penting dalam mengukur profitabilitas nilai pelanggan [2].

Dalam menguji model RFM ini, peneliti menggunakan data histori Perusahaan Distributor PT.Cendana Perdana Perkasa di Pekanbaru yang bergerak dibidang penjualan Kerupuk Nixxa dan Senna. Perusahaan ini belum pernah melakukan pengukuran perilaku pembelian pelanggan dengan model RFM ini, karena sebelumnya hanya melihat transaksi keseluruhan. Pengukuran ini menggunakan klasterisasi yang menganggap bahwa jika pelanggan berada pada klaster yang sama maka dianggap perilaku pelanggan sama dengan pelanggan yang ada di dalam klaster tersebut. Klasterisasi tersebut akan dinamakan dengan segmentasi. Klasterisasi pada penelitian ini menggunakan metode clustering K-Means yang membagi data menjadi beberapa segmen. Algoritma K-Mean clustering telah banyak digunakan pada penelitian sebelumnya dalam melakukan segmentasi pelanggan seperti pada penelitian [7].

2. METODE PENELITIAN

Penelitian ini menggunakan beberapa metode yang bisa dilakukan dalam menyelesaikan permasalahan yang terdiri dari beberapa tahapan :

2.1 Model RFM

Model ini terdiri dari 3 variabel yaitu :

- Recency yaitu tanggal terakhir transaksi yang dilakukan pelanggan pada perusahaan selama periode analisis.
- Frekuensi adalah jumlah transaksi yang dilakukan oleh pelanggan pada perusahaan selama periode yang dianalisis
- Monetary yaitu jumlah uang yang dikeluarkan pelanggan untuk perusahaan selama periode analisis.

Data perusahaan berupa nilai transaksi, jumlah transaksi dan jumlah uang akan dikelompokkan berdasarkan model R, F dan M. Data perusahaan yang diambil dengan periode analisis satu tahun mulai dari 01 Agustus 2016 sampai dengan 31 Juli 2017.

2.2 Metode Min-Max

Penelitian ini menggunakan normalisasi dengan metode Min-Max. Normalisasi diperlukan agar data R, F dan M tidak terlalu jauh sehingga digunakan range antara 0 sampai 1 dengan metode Min-Max. Range 0-1 digunakan agar nilai normalisasi pada penelitian ini memiliki nilai antara 0 dan 1. Nilai $R(ci)$, dan $F(ci)$ dan $M(ci)$ merupakan hasil normalisasi yang akan didefinisikan sebagai berikut (Khajvand *et al.*, 2011) :

$$R(ci) = \frac{o_i^R - \min^R}{\max^R - \min^R} (\text{newmax}^R - \text{newmin}^R) + \text{newmin}^R \quad (1)$$

$$F(ci) = \frac{o_i^F - \min^F}{\max^F - \min^F} (\text{newmax}^F - \text{newmin}^F) + \text{newmin}^F \quad (2)$$

$$M(ci) = \frac{o_i^M - \min^M}{\max^M - \min^M} (\text{newmax}^M - \text{newmin}^M) + \text{newmin}^M \quad (3)$$

O_i^R , O_i^F dan O_i^M merupakan data mentah yang akan dinormalisasikan pada masing-masing variabel R, F dan M. Sedangkan \min^R , \min^F dan \min^M adalah nilai minimal dari data mentah pada masing-masing R, F dan M. Begitu juga \max^R , \max^F dan \max^M yang merupakan maksimal dari data mentah pada masing-masing R, F dan M. Pada newmax dan newmin baik itu untuk R, F dan M menggunakan range antara 0 dan 1, dimana 0 adalah nilai minimal dan 1 adalah nilai maksimal.

2.3 Silhouette Index

Dalam menentukan jumlah Cluster yang optimal bisa divalidasi dengan menggunakan *Silhouette Index*. Tujuan dari teknik validasi *cluster* adalah untuk mengevaluasi hasil *clustering*, hasil evaluasi ini dapat digunakan untuk menentukan jumlah *cluster* pada dataset. Salah satu teknik validasi

cluster adalah *Silhouette Index* yaitu metode penafsiran untuk validasi kluster pada objek-objek. Teknik ini memberikan representasi grafis singkat tentang seberapa baik setiap objek terletak dalam *clusternya*. *Silhouette Index* dikembangkan pertama kali oleh Rousseeuw pada tahun 1986. Nilai *Silhouette Index* (SI) didapatkan dengan persamaan sebagai berikut :

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (4)$$

Dimana $s(i)$ merupakan nilai *Silhouette Index*, $a(i)$ merupakan rata-rata jarak antara titik i dengan seluruh titik dalam A (*cluster* dimana titik i berada), $b(i)$ adalah rata-rata jarak antara titik i ke seluruh titik dalam *cluster* selain A .

2.4 K-Means

Metode K-Means adalah metode yang dikategorikan ke dalam metode clustering partisi [3]. Algoritma ini paling banyak digunakan dalam CRM dan pemasaran [3]. Hal ini disebabkan karena algoritma ini memiliki kesederhanaan dan kemudahan dalam menggunakannya dan pengguna bisa menentukan sendiri jumlah cluster nya. Tahapan dalam metode K-means adalah sebagai berikut :

- [1] Tentukan jumlah kluster
- [2] Pilih centroid awal secara acak sesuai jumlah kluster
- [3] Hitung jarak data ke centroid dengan rumus *euclidean distance*.

$$d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

- [4] Perbarui centroid dengan menghitung nilai rata-rata nilai pada masing-masing kluster
- [5] Kembali ke tahapan ke 3 jika masih terdapat data yang berpindah kluster atau perubahan nilai centroid

Pada tahap 1 dan 2 dari tahapan K-Means, jumlah kluster ditentukan dengan metode dunn index dan *silhouette coefficient* yang merupakan metode validasi dalam menentukan jumlah kluster yang terbaik.

2.5 Pemetaan Pelanggan

Pemetaan pelanggan dilakukan untuk menganalisis perilaku pelanggan berdasarkan segmennya dengan melihat nilai R, F dan M. Didalam penelitian [3] mengelompokkan 8 kelompok pelanggan dengan tambahan satu atribut yaitu L (Length). Penelitian ini hanya menggunakan atribut R, F dan M. Simbol (↑) merepresentasikan nilai kelompok di atas rata-rata keseluruhan. Sementara simbol (↓) merepresentasikan nilai kelompok di bawah rata-rata keseluruhan. Tiap segmen akan dihitung nilai R, F, dan M, jika nilai R, F, dan M menghasilkan nilai di atas rata-rata dari keseluruhan nilai R, F dan M maka akan dilambangkan dengan simbol (↑) artinya bahwa segmen tersebut bagus. Simbol (↑) ini dikatakan bagus untuk atribut F dan M saja. Namun nilai R dengan simbol (↓) inilah yang dikatakan bagus karena semakin kecil nilai recency ini berarti bahwa segmen ini baru-baru ini melakukan transaksi.

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini sebanyak 12 bulan yang berjumlah 5069 record dan data yang bisa dianalisis berjumlah 4529 record dikarenakan data 540 tersebut terdapat data minus hasil penjualan dari pengembalian outlet ke perusahaan dengan periode analisis 1 Agustus 2016 sampai 31 Juli 2017. Dari data tersebut maka akan dipilih data yang memiliki atribut R, F dan M. Selanjutnya akan dicari nilai minimal dan maksimal dari masing-masing variabel R, F dan M. Tabel 1 merupakan hasil penilaian minimal dan maksimal dari data pelanggan berdasarkan model RFM.

Tabel 1. Hasil Penilaian Min-Max Pada Model RFM

Variabel RFM	Min	Max	Rata-Rata
Recency	0	539	88
Frekuensi	1	165	15
Monetary	Rp 35.937	Rp 231.838.791	Rp. 5.321.117

Setelah mengetahui nilai minimal dan maksimal dari masing masing variabel R,F dan M maka persamaan (1), (2) dan (3) akan digunakan untuk mendapatkan normalisasi data setiap pelanggan sesuai dengan variabel R, F dan M. Tabel 2 merupakan potongan data dari hasil normalisasi RFM. Nilai $R(ci)$, $F(ci)$ dan $M(ci)$ menggunakan lambang R_N , F_N dan M_N dimana N adalah normalisasi pada Tabel 2.

Tabel 2. Potongan Data Normalisasi Pembobotan RFM

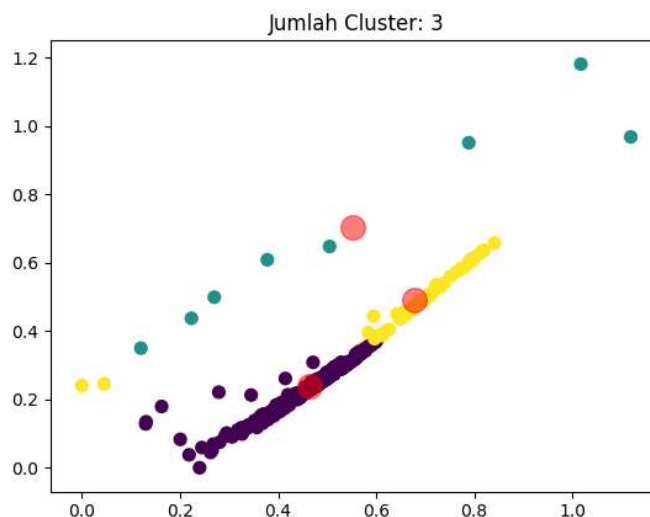
Nama Pelanggan	R	R_N	F	F_N	M	M_N
999 Swalayan	4	0,007	86	0,518	24072962	0,104
Anton Lucky	41	0,076	49	0,293	13302056	0,057
Besta Jaya	53	0,098	15	0,085	1820808	0,008
Belimbing Jaya	102	0,189	5	0,024	765930	0,003
Erina Lestari	6	0,011	40	0,238	5416398	0,023
Greens Mart	252	0,468	2	0,006	319440	0,001
Mama Sukses Abadi	5	0,009	34	0,201	6956678	0,030
Tiga "M" Sukses	62	0,115	26	0,152	4287037	0,018
Depo Cendana	10	0,019	34	0,201	55643431	0,240
Family Swalayan	342	0,635	3	0,012	479160	0,002
Giant Nangka Pku	2	0,004	143	0,866	188199082	0,812
Giant Metropolitan City	31	0,058	51	0,305	64171941	0,277
Giant SPM Arengka	31	0,057	12	0,067	7404431	0,032
Hawaii Pasar Swalayan	53	0,098	21	0,122	4062284	0,017
Hawaii Swalayan Rumbai	7	0,013	88	0,530	14137750	0,061
Kandis Swalayan	326	0,605	3	0,0122	479160	0,002
Kd. Aas	11	0,020	24	0,140	2715240	0,012
Kd. Ajo	14	0,026	11	0,061	1042470	0,004
Kd. Akbar	18	0,033	7	0,037	503580	0,002
Kd. Alinar 2	62	0,115	8	0,043	383328	0,001

Setelah semua data dinormalisasikan maka selanjutnya akan dilakukan penentuan jumlah kluster dalam penentuan segmen pelanggan dengan melihat nilai Silhouette Index dengan menggunakan Program *Phyton*. Nilai SI yang tertinggi pada suatu kluster yang akan dianalisis selanjutnya.

Tabel 3. Hasil Penentuan Kluster dengan Silhouette Index

Jumlah kluster	Silhouette Index
2	0.588
3	0.589
4	0.466
5	0.497
6	0.498
7	0.472
8	0.419
9	0.428
10	0.405

Berdasarkan Tabel 3 maka dapat dilihat bahwa nilai SI tertinggi berada pada 3 kluster. Meskipun nilainya tidak terlalu jauh berbeda dengan 2 kluster. Untuk melihat sebaran data pada 3 kluster dapat dilihat pada Gambar 1. Pada Gambar 1, ada 3 kluster terdapat sebaran dua data pada kluster 2 (warna kuning) yang datanya berjauhan dengan tumpukan data pada kluster tersebut. Warna kuning pada Gambar 1 merupakan data yang berada pada kluster 1, warna ungu biru pada kluster 2 dan warna biru pada kluster 3.



Gambar 1. Plot 3 Kluster

Berdasarkan Tabel 3 dan Gambar 1, pelanggan akan dibagi menjadi 3 Segmen sesuai dengan jumlah kluster. Pembagian pelanggan ini menggunakan teknik klasterisasi K-Means. Dari hasil pengolahan menggunakan teknik K-Means maka didapatkan nilai titik tengah kluster yang dapat dilihat pada Tabel 4.

Tabel 4. Nilai Titik Tengah (Centroid) RFM

Atribut	Segmen 1	Segmen 2	Segmen 3
R_N	0.019	0.453	0.082
F_N	0.624	0.023	0.078
M_N	0.270	0.009	0.013

Berdasarkan Tabel 2 dengan menggunakan teknik klasterisasi K-Means maka akan dihasilkan jumlah Segmen per pelanggan yang dapat dilihat pada Tabel 5.

Tabel 5. Nilai Klaster Per Pelanggan

Nama Pelanggan	Segmen
999 Swalayan	1
Anton Lucky	2
Besta Jaya	2
Belimbing Jaya	2
Erina Lestari	2
Greens Mart	1
Mama Sukses Abadi	2
Tiga "M" Sukses	2
Depo Cendana	2
Family Swalayan	1

Giant Nangka Pku	3
Giant Metropolitan City	2
Giant SPM Arengka	2
Hawaii Pasar Swalayan	2
Hawaii Swalayan Rumbai	1
Kandis Swalayan	1
Kd. Aas	2
Kd. Ajo	2
Kd. Akbar	2
Kd. Alinar 2	2

Tabel 6. Jumlah Data Pelanggan 3 Segmen

Jumlah Segmen	Jumlah Data	Simbol Rank RFM	Rank Segmen Terbaik
1	57	R↑F↓M↓	3
2	257	R↓F↑M↓	2
3	8	R↓F↑M↑	1

Langkah selanjutnya adalah menganalisis perilaku pembelian pelanggan berdasarkan klusternya masing-masing. Perilaku pembelian pelanggan berdasarkan kluster sebagai berikut :

- 1) Segmen 1 yaitu *recency* tinggi artinya bahwa pelanggan pada segmen ini sudah lama tidak melakukan transaksi, *frequency* nya rendah artinya ini merupakan pelanggan yang tergolong melakukan transaksi dengan perusahaan rata-rata kecil serta *monetary* rendah yang artinya jumlah uang yang dibelanjakan rata-rata juga kecil. Segmen ini bisa dikatakan pelanggan yang tidak jelas atau belum pasti akan menjadi potensial.
- 2) Segmen 2 yaitu *recency* yang rendah artinya baru-baru ini melakukan transaksi, *frequency* nya tinggi artinya ini merupakan pelanggan yang tergolong sering melakukan transaksi dengan perusahaan dan *monetary* rendah yang artinya jumlah uang yang dibelanjakan tidak begitu besar. Segmen ini adalah juga bisa dikatakan pelanggan yang potensial namun perusahaan harus bekerja keras bagaimana mempertahankan pelanggan tersebut dan meningkatkan jumlah nominal uang yang dibelanjakan atau nilai *monetary* nya.
- 3) Segmen 3 yaitu pelanggan yang memiliki *recency* yang rendah artinya bahwa pelanggan pada segmen ini baru melakukan pembelian, *frequency* dan *monetary* tinggi artinya bahwa segmen ini merupakan segmen yang jumlah pembelian dan uangnya tinggi sehingga segmen ini merupakan pelanggan yang potensial untuk ke depan. Pelanggan pada segmen ini sangat perlu dipertahankan.

4. KESIMPULAN

Penelitian ini menghasilkan nilai 3 kluster yang terbentuk dengan menggunakan metode klasterisasi K-means dengan nilai *Silhouette* 0,589 dengan bahasa pemrograman *Phyton*. Berdasarkan hasil pengklasterisasi tersebut maka dapat disimpulkan bahwa perilaku pembelian pelanggan memiliki pola yang berbeda-beda. Kluster yang memiliki perilaku pelanggan yang baik berada pada kluster 3 dan kluster 2 (meskipun kluster ini perlu dilakukan usaha dalam meningkatkan nilai uang pelanggan bagi perusahaan). Kluster yang baik perlu dipertahankan agar pelanggan tersebut tetap memiliki nilai R yang rendah, F yang tinggi dan M yang tinggi dan juga bisa dilakukan strategi dalam meningkatkan nilai kluster tersebut.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada keluarga, teman dan institusi yang telah memberikan dukungan baik itu moril dan material dalam terselesainya penelitian ini.

BAHAN REFERENSI

- [1] Buttle, F. , 2008, *Customer Relation Management, Concept and Technologies, Second Edition*, Elsevier Butterworth-Heinemann. doi: 10.1007/s13398-014-0173-7.2.
- [2] He, X. and Li, C. ,2016, 'The Research and Application of Customer Segmentation on E-commerce Websites'. doi: 10.1109/ICDH.2016.47.
- [3] Kandeil, D. A., Saad, A. A. and Youssef, S. M. , 2014, 'A two-phase clustering analysis for B2B customer segmentation', *Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014*, pp. 221–228. doi: 10.1109/INCoS.2014.49.
- [4] Khajvand, M., Zolfaghar, K., Ashoori, S. and Alizadeh, S. ,2011, 'Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study', *Procedia Computer Science*. Elsevier, 3, pp. 57–63. doi: 10.1016/j.procs.2010.12.011.
- [5] Kim, S.-Y., Jung, T.-S., Suh, E.-H. and Hwang, H.-S. ,2006, 'Customer segmentation and strategy development based on customer lifetime value: A case study', *Expert Systems with Applications*, 31(1), pp. 101–107. doi: 10.1016/j.eswa.2005.09.004.
- [6] Ngai, E. W. T., Xiu, L. and Chau, D. C. K. ,2009, 'Application of data mining techniques in customer relationship management: A literature review and classification', *Expert Systems with Applications*. Elsevier Ltd, 36(2 PART 2), pp. 2592–2602. doi: 10.1016/j.eswa.2008.02.021.
- [7] Panuš, J., Jonášová, H., Kantorová, K., Doležalová, M. and Hořáčková, K. ,2016, 'Customer segmentation utilization for differentiated approach', *IDT 2016 - Proceedings of the International Conference on Information and Digital Technologies 2016*, pp. 227–233. doi: 10.1109/DT.2016.7557178.
- [8] Nasution, Muhammad Irwan Padli, 2014, *Keunggulan Kompetitif dengan Teknologi Informasi. Jurnal Elektronik*
- [9] A. Hughes, *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program*. McGraw-Hill Companies, Incorporated, ,2006, [Online]. Available: <http://books.google.com.eg/books?id=Ws cBzKDk8MC>